

# Classification of Multiple Electromagnetic Interference Events in High Voltage Power Plant

Imene Mitiche, Gordon Morison, Alan Nesbitt, Michael Hughes-Narborough  
*Glasgow Caledonian University*  
Glasgow, UK

Brian G. Stewart  
*University of Strathclyde*  
Glasgow, UK

Philip Boreham  
*Doble Engineering*  
Bere Regis, UK

**Abstract**—This paper addresses condition assessment of electrical assets contained in high voltage power plants. Our work introduces a novel analysis approach of multiple event signals related to faults, and which are measured using Electro-Magnetic Interference method. The proposed method transfers the expert's knowledge on events presence in the signals to an intelligent system which could potentially be used for automatic EMI diagnosis. Cyclic spectrum analysis is used as feature extraction to efficiently extract the repetitive rate and the dynamic discharge level of the events, and multi-class support vector machine is adopted for their classification. This first and novel method achieved successful results which may have potential implications on developing a framework for automatic diagnosis tool of EMI events.

**Keywords**—*electrical condition assessment, electro-magnetic interference, partial discharge, expert system, pattern recognition*

## I. INTRODUCTION

The Electromagnetic Interference (EMI) method is one way to identify faults in insulation and conduction media within high voltage assets including motors, cables, transformers and associated auxiliary equipment [1]. This technique captures the EMI radiation or Radio Frequency (RF) conduction that arises from fault related discharge such as corona, arcing, sparking and Partial Discharge (PD). The latter is normally an indication of insulation degradation, broken conductors, or loose connections [2]. Thus, identification of these faults using EMI methods may be helpful to obtain information on a high voltage asset condition permitting further decisions and actions to be taken on the operating asset's maintenance to avoid any catastrophic failure. This has the advantage of reduced replacement cost and system down time, enhanced safety, maximized return of investment and business profit. EMI data is collected during the operation of the asset. Usually, EMI experts analyse the audio spectrum of the captured data [2] [3]. This method tends to be time consuming and it is not practical for continuous monitoring. Furthermore, the availability of

such experts is scarce. This paper aims to develop a software system that captures and utilises expert knowledge on EMI condition monitoring. This type of approach for EMI condition monitoring is relatively new and only a few publications exist at present [4]. Our approach exploits machine learning techniques which involve feature extraction and classification of high voltage field data that represents the different discharges. It is essential to retrieve the unique fingerprint of each discharge type for successful classification. However, it is often challenging for complex signals. Suitable feature extraction techniques should be employed according to the data nature. The data in this work was obtained from sites that contain operating rotating machines. It was reported in [5] that mechanical or electrical faults in rotating machines generate a cyclic signature that could be present in the captured data signals. This signature determines periodic statistical characteristics within the cyclostationary signal. Cyclic Spectral Analysis (CSA) was introduced in [6] as a tool for mechanical fault detection by obtaining information on the statistical behavior of fault signals which happen to exhibit modulated or repeated transients. Thus, this provides the motivation to investigate for the first time the application of CSA to EMI signals. The aim is to extract a relevant pattern associated with each fault in feature extraction stage of the classification process. A feature extraction technique for time series data called 1D Local Binary Pattern (LBP) was proposed in [7] and is utilised in this work. Firstly, CSA is applied to extract the cyclostationary behavior of EMI signals producing an alpha-domain on which 1D LBP is calculated for feature reduction and improved classification purposes. A Multi-Class Support Vector Machine (MCSVM) classifier is employed as a classification technique to classify the multiple discharges. The next section describes the method of EMI measurement. The paper is structured as follows. Section III provides the definition of feature extraction and classification algorithms. Section IV describes the experimental method followed by results and conclusions.

## II. EMI DATA MEASUREMENT

EMI signals are produced due to different fault types in generators, cables, motors and related auxiliary equipment [8]. The electromagnetic pulses spread from any discharge event caused by a fault in all directions. These signals can be sensed and measured over a large frequency range that varies from 10's of kHz up to 100MHz in an Isolated Phase Bus (IPB), for example. EMI energy can be either radiated or conducted. The conducted energy is detected and is measured using a split-core High Frequency Transformer (HFCT), located on the neutral connection of an asset. In order to establish measurement compatibility and interpretation of results between EMI measurement instruments, the EMI measurement is performed following the CISPR16-1-1 standard. The emitted EMI signals may represent mechanical or electrical defect such as loose or broken stator and rotor bars, winding insulation degradation etc. The fault level of degradation, location and severity can be acquired by the EMI method some way before identification using traditional methods. This is because the resulting EMI frequency spectrum is unique for each fault type and each physical location in an electrical system and is therefore more suitable for system diagnostics application. Thus, the EMI analysis technique is able to differentiate between different fault types and to assess the condition of a generator as well as the condition of adjacent auxiliary assets. Various defects have been identified by means of the EMI technique within cables, exciters, transformers, IPB and machine stator windings [9].

## III. CLASSIFICATION ALGORITHMS

This section provides a brief description of the overall process for EMI discharge classification and the mathematical theory of the employed algorithms. Fig. 1. illustrates the methods used for on-site EMI data measurement and pattern recognition including feature extraction, using cyclic spectrum and 1D LBP, and the MCSVM classifier. A detailed explanation for each stage is provided as follows.

### A. Cyclic Spectrum Analysis

A signal is said to be cyclostationary at an order 'p' if it exhibits periodic statistical characteristics at the order 'p' [10].

Let  $x[n]$ ,  $n=0, 1, \dots, M-1$  be the signal of interest captured at a sampling frequency  $F_s$ . The Short Time Fourier Transform (STFT) of  $x[n]$  across a time window with length of  $N_w/F_s$  is written as:

$$X_{STFT}(i, f_k) = \sum_{m=0}^{N_w-1} x[iR + m] \cdot w[m] \cdot e^{-j2\pi m f_k / F_s} \quad (1)$$

where  $f_k = k\Delta f$ ,  $k=0, \dots, N_w-1$ ,  $R$  is a time shift between successive windows at time index  $i$ , and  $w[m]$  is a symmetric window. The instantaneous autocorrelation function of  $x[n]$  is defined as:

$$R_x(t_n, \tau) = E\{x(t_n - \tau) \cdot x^*(t_n)\} \quad (2)$$

where  $E$  is the ensemble average,  $(.)^*$  is the complex conjugate,  $t_n = n/F_s$  are time instants at which the signal samples were acquired, and  $\tau$  is a time lag.  $R_x$  of a cyclostationary signal is

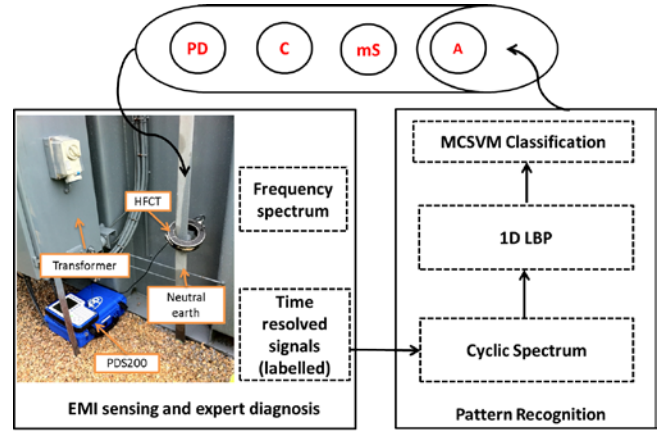


Fig. 1. EMI measurement and classification process of Partial Discharge (PD), Corona (C), micro Sparking (mS) and Arcing (A).

periodic, and this exhibits the presence of a repetitive statistical comportment in the signal. In fault signals, the occurrence of a series of impulses or periodic modulation leads to such behaviour. Next, the Spectral Correlation (SC) introduced as the double Discrete Fourier Transform (DFT) of  $R_x$  with respect to time  $t$  and the time lag  $\tau$  which is defined as:

$$S_x(\alpha, f) = \lim_{N \rightarrow \infty} \frac{1}{(2N+1)f_s} \sum_{n=-N}^N \sum_{m=-\infty}^{\infty} R_x(t_n, \tau_m) e^{-j2\pi n \frac{\alpha}{f_s}} e^{-j2\pi m \frac{f}{f_s}} \quad (3)$$

where  $\tau_m = m/F_s$ . SC of a cyclostationary signal reveals the characteristics of a pattern that is continuous in frequency  $f$  and discrete in a cyclic frequency  $\alpha$ , given that:

$$S_x(\alpha, f) = \begin{cases} S_x^k(f), & \alpha = k/T \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where  $S_x^k(f)$ ,  $k=0, \pm 1, \dots$  is the cyclic spectrum at  $k$ . This results in a bispectrum that contains a number of cyclic spectrums at cyclic frequencies  $\alpha = k/T$ . The special case of  $\alpha=0$  is basically the classic power spectrum density. The cyclic frequency  $\alpha$  and the frequency  $f$  can also be seen as the modulation and carrier frequencies respectively, of a modulated signal. The SC or cyclic spectral density was demonstrated in [10] to be a suitable tool for fault condition monitoring because it has the ability to retrieve and transfer the characteristics of complex signals into a 2-D map that illustrates the modulation frequencies which are potentially related to fault frequencies. Fig. 2 shows the cyclic spectral density of example PD and micro sparking EMI signals. It is observed that each EMI signal type has a unique and different signature. In order to remove the redundant information and reduce the feature dimension, 1D LBP of the  $\alpha$  spectrum is calculated to produce the final feature vector.

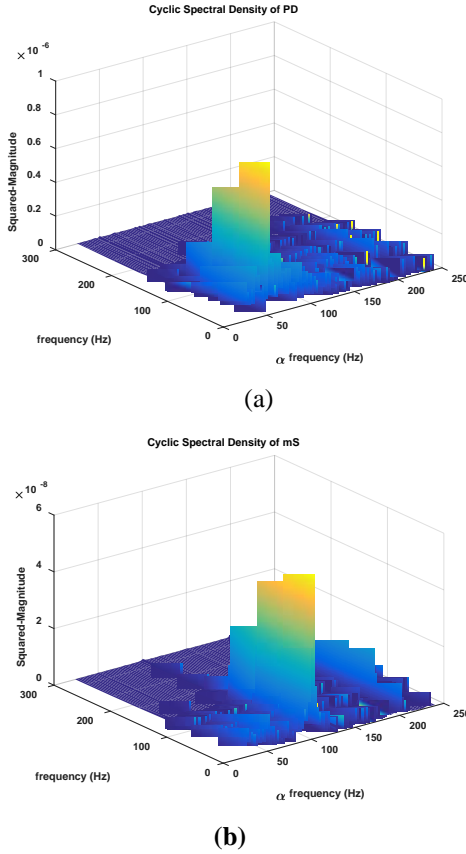


Fig. 3. Cyclic Spectral Density of (a) PD (b) micro Sparking (mS).

### B. 1D Local Binary Pattern

This method is inspired by an image feature extraction technique known as 2D LBP. The aim of 1D LBP is to process a single vector of data samples, and the idea is to map each data sample to an LBP code by thresholding against its neighboring samples. The 1D LBP theory is described as follows. Let  $x[n]$ ,  $n=0, 1, \dots, N$  be the data vector and  $P$  is the number of neighboring samples. The 1D LBP of  $x[n]$  is formulated as:

$$LBP = \sum_{r=0}^{\frac{P}{2}-1} \left\{ \begin{aligned} &S \left[ x[n+r-\frac{P}{2}] - x[n] \right] 2^r \\ &+ S \left[ [x[n+r+1] - x[n]] \right] 2^{r+\frac{P}{2}} \end{aligned} \right\},$$

$$S.t. \quad S[y] = \begin{cases} 1 & \text{if } y \geq 0 \\ 0 & \text{if } y < 0 \end{cases} \quad (5)$$

An example of this concept is presented in Fig. 3 for 1D LBP calculated over a time series for  $P=8$ . This LBP code is standard and contains  $2^P$  codes. Here, uniform LBP is employed because it results in only 59 codes. This has the benefit to reduce the feature vector dimension while providing robust and non-redundant information. An LBP is considered uniform if the binary code contains a maximum of two 01 or 10 transitions. A full explanation of uniform LBP is provided in [11]. The LBP values describe the variation in the data. The difference between the center sample and neighboring samples

is close to zero when the data is steady or slowly varying. In contrast, this difference is larger when the data is rich in narrow pulses.

### C. Support Vector Machine

SVM is a binary and a regression classifier. The binary classifier separates two different classes with a hyperplane which can be obtained linearly or through a kernel function e.g. polynomial or Radial Basis Function (RBF). The advantage of using SVM is the ability to handle large features while achieving high classification accuracy [11]. Let inputs to the SVM classifier be  $x_i$  and  $y_i$ ;  $i=1, 2, \dots, L$  which represent the data and their relative labels respectively. Assuming that the data samples are among two classes +1 and -1, they are separated by a linear hyperplane defined as:  $f(x) = w \cdot x + b = 0$ , where  $b$  is a scalar and  $w$  is a  $P$ -dimensional vector that defines the position of the hyperplane. A hyperplane for the first and second class is defined as  $w \cdot x + b = 1$  and  $w \cdot x + b = -1$  respectively. The hyperplanes are obtained by solving the optimization problem presented in the following equation.

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^M \zeta_i, \quad (6)$$

$$S.t. \quad \begin{cases} y_i (w^T \cdot x_i + b) \geq 1 - \zeta_i \\ \zeta_i \geq 0, \quad i = 1, \dots, M \end{cases}$$

where  $C$  is the error penalty that determines the trade-off between classification error and margin maximization in the training phase, and  $\zeta_i$  is the noise slack which represents the distance between the margin and the samples that are in error by crossing the margin line. The Lagrange method is adopted to tackle the optimization problem in (9). Further explanation on how this is achieved is provided in [12]. The Lagrange strategy is explained in detail in [13]. In this work, the binary SVM classifier is exploited in an MCSVM method for the aim of multiple EMI fault events classification. Specifically, a one-against-one approach is implemented [14] which utilizes a number of models comprising of SVM pairs. Each model is trained on two classes.

## IV. EXPERIMENTAL METHOD, RESULTS AND CONCLUSIONS

### A. EMI data acquisition

The EMI data analysed in this work was measured during apparatus operation using a PDS200 device which operates under the CISPR16 standard for EMC type filtering. This device provides a frequency spectrum of the measured data and the possibility to retrieve time resolved signals sampled at 24kHz. “EMI experts” examine and assess the audio signals and identify the fault or EMI event type in each signal. From expert analysis and labelling, the data set studied in this paper contains four main EMI types: PD, corona, micro sparking and arcing. The signals were measured in different power sites and from different apparatus. The corona signal originates from the neutral earth cable of a generator. PD signals were identified in eleven different sites at generator, IPB and transformer. Arcing and micro sparking were identified in two sites at a generator. Each measured signal is 10 seconds long

and contains 500 cycles. Three sites among the ten sites were found to contain two or more EMI classes.

### B. Classification method and Results

First, each EMI signal is segmented with a non-overlapping window of 10k samples for easy and fast computation of the CSA. Cyclic spectral density is obtained for each segment. A summation over the frequency domain is performed in order to obtain the alpha domain pattern which is considered as the time series. 1D LBP is then calculated to extract the relevant features from each CSA signature. Each feature vector was labelled with respect to the identified EMI event or fault by "EMI experts", and implemented to the MCSVM algorithm. The classification is performed using a standard strategy known as 10 fold cross validation. In this strategy a random 90% of the data along with the labels is used for training and the remaining unseen 10% without labels is left for testing. This is performed ten times, each time with a different random training/testing batch. The whole process is repeated for segments of 20, 30, 40 and 50k samples to determine the nature of accuracy improvement and consistency as a function of data segment size to assist in future optimization methods for such a technique.

The proposed algorithm's performance was evaluated in terms of classification accuracy which is obtained by comparing the classifier's output label and the true label of the testing data set. The considered accuracy is the average result of the ten folds classification. Results for different segment sizes are presented in table 1. It is interesting to point out that the classification accuracy is directly proportional to the segment size of the time series signals. The maximum classification accuracy is achieved for 40k samples/segments and appears not to improve further for higher number of samples per segment. The explanation to this could be that the CSA captures significant cyclostationary information for a particular number of cycles in a signal.

TABLE I. CLASSIFICATION ACCURACY RESULTS

# samples	10k	20k	30k	40k	50k
accuracy (%)	85	92	94	96	96

### C. Conclusions

This work aims to develop a condition monitoring system for high voltage assets through a pattern recognition approach based on an expert system, where specifically the proposed algorithms target real world EMI signals. Robust feature extraction techniques are employed for successful performance of the pattern recognition. High classification accuracy of four

EMI faults (PD, corona, arcing and micro sparking) was achieved. The significant contribution of this work is mainly the successful multiple fault identification in EMI signals by means of CSA signature, and secondly the successful classification of signals originating from different apparatus and multiple sites. The proposed algorithm opens instrumentation possibilities for EMI fault diagnosis in high voltage power plants.

### REFERENCES

- [1] J. E. Timperley and J. M. Vallejo, "Condition assessment of electrical apparatus with EMI diagnostics," IEEE Trans. Indus. App., vol. 53, pp. 693-699, January 2017.
- [2] J. E. Timperley, "Detection of insulation deterioration through electrical spectrum analysis," in 1983 EIC 6<sup>th</sup> Elec./Electro Insu. Conf., Chicago, USA, pp.60-64, October 1983.
- [3] J. E. Timperley, "Audio spectrum analysis of EMI patterns," in 2007 Electrical Insulation Conference and Electrical Manufacturing Expo, Nashville, USA, pp.39-41, October 2007.
- [4] I. Mitiche, G. Morison, M. Hughes-Narborough, A. Nesbitt, P. Boreham, B.G. Stewart, "Classification of partial discharge signals by combining adaptive local iterative filtering and entropy features", IEEE CEIDP, Fort Worth, USA, pp.335-338, October 2017.
- [5] J. Antoni, G. Xin and N. Hamzaoui, "Fast computation of the spectral correlation,," Mech. Sys. And Sig. Proc., vol. 92, pp.248-277, 2017.
- [6] A. Mahvash and A. A. Lakis, "Application of Cyclic Spectral Analysis in Diagnosis of Bearing Faults in Complex Machinery," Trib. Trans., vol. 58, pp. 1151-1158, November 2017.
- [7] P. McCool, N. Chatlani, L. Petropoulakis, J. J. Soraghan, R. Menon and H. Lakany, "1-D local binary patterns for onset detection of myoelectric signals, " in 20<sup>th</sup> Euro. Sig. Proc. Conf., Bucharest, Romania, pp. 499-503, August 2012.
- [8] J.E. Timperley and J.M. Vallejo, "Condition assessment of electrical apparatus with EMI diagnostics", In IEEE Petr. & Chem. Indu. Comm. Conf., Houston, USA, pp.1-8, October 2015.
- [9] J.E. Timperley, J.M. Vallejo and A. Nesbitt, "Trending of EMI data over years and overnight", In IEEE Elec. Insu. Conf., Philadelphia, USA, pp.176-179, June 2014.
- [10] M. Kedadouch, M. Thomas and A. Tahan, "Cyclostationarity applied to acoustic emission and development of A New Indicator for Monitoring Bearing Defects , " Mech. & Indu., vol. 15, pp. 467-476, September 2014.
- [11] M. Topi, O. Timo, P. Matti and S. Maricor, "Robust texture classification by subsets of Local Binary Patterns, " in Proc. Of the 15<sup>th</sup> Intl. Conf. on Patt. Rec., Barcelona, Spain, pp. 971-987, September 2000.
- [12] C. Bishop, Pattern recognition and machine learning, Cambridge, UK: Springer, 2006.
- [13] A. Widodo, B.-S. Yang, "Support Vector machine in machine condition monitoring and fault diagnosis", Mech. Sys. & Sig. Proc., vol. 21, pp. 2560-2574, August 2007.
- [14] L. Bottou, C. Cortes, J.S. Denker, H. Drucker, I. Guyon, L. Jackel, Y. LeCun, U.A Muller, E. Sackinger, P. Simard and V. Vapnik, "Comparison of classifier methods: a case study in handwriting digit recognition", in Int. Conf. on Patter. Recog., Jerusalem, Israel, pp. 77-87, October 1994.